INFLUENCE OF CHOICE OF TIME PERIOD ON GLOBAL SURFACE TEMPERATURE TREND ESTIMATES

BY BRANT LIEBMANN, RANDALL M. DOLE, CHARLES JONES, ILEANA BLADÉ, AND DAVE ALLURED

Annual global and continental temperature changes for the period 1850–2009 are examined for varying time intervals and year ranges.

Assessments of variations in global average surface temperature have generated substantial interest and controversy within and beyond the climate science community. One of the most interesting and controversial aspects of the observed changes is the slope, or trend, of global mean temperatures.

An important consideration when analyzing trends of any kind is the choice of start and end dates of the time series. Trend estimates can change dramatically by including or excluding a few years at either end of the time series, particularly when computing from relatively short time series. Studies typically employ either the longest available record for a given dataset or a shorter period that is common to two or more variables of interest, with the intent of identifying temporal relationships between variables and making attribution statements for the observed trends. Sometimes, however, the choice of period may be somewhat arbitrary.

The present study provides a simple method to assess the influence of the choice of beginning and end years on trend estimates. While the method can be applied to many kinds of observations, this note focuses mainly on global surface temperature records as used in the recent Intergovernmental Panel on Climate Change Fourth Assessment Report (Solomon et al. 2007). The ultimate purpose of this study is to estimate time scales over which the trend of global average surface temperature appears to be robust. We also examine recent cooling trends that have received widespread attention in various parts of the press and other public media. We find that such short-term trends occur often, lack statistical significance, and should therefore not be interpreted as indicative of longer-term trends.

DATA. The primary datasets used in this study are the historical average surface temperature time series, from 1850 to 2009, produced by the University of East Anglia (UEA) Climate Research Unit, in conjunction with the Hadley Centre of the Met Office,

AFFILIATIONS: LIEBMANN AND ALLURED—University of Colorado, Boulder, Colorado; DOLE—NOAA/Earth System Research Laboratory, Boulder, Colorado; JONES—Institute for Computational Earth System Science, University of California, Santa Barbara, Santa Barbara, California; BLADÉ—Departament d'Astronomia i Meteorologia, Universitat de Barcelona, Barcelona, Spain
CORRESPONDING AUTHOR: Brant Liebmann, NOAA/ESRL/PSD Climate Diagnostics Branch, R/PSD1, 325 Broadway, Boulder, CO 80305–3328
E-mail: brant.liebmann@noaa.gov

The abstract for this article can be found in this issue, following the table of contents.
DOI:10.1175/2010BAMS3030.1

In final form 26 April 2010

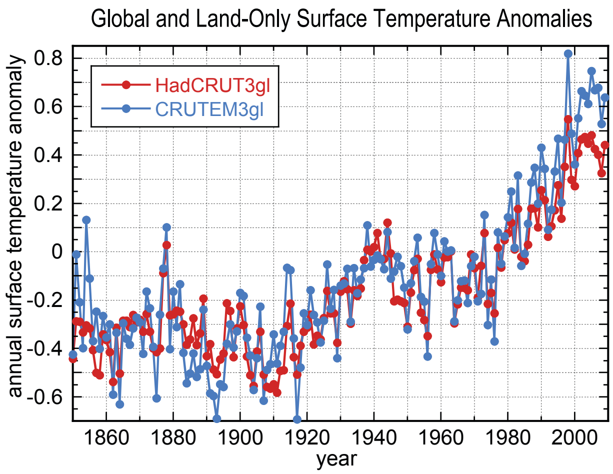

Fig. 1. Annual mean temperature anomalies for the period 1850–2009 (°C), expressed as departures from 1961–90 average. Red curve represents global anomalies (HadCRUT3) and blue curve represents land-only anomalies (CRUTEM3).

for both global (HadCRUT3) and land surface only (CRUTEM3) temperature (Brohan et al. 2006; Jones 1994; Jones and Moburg 2003; Jones et al. 1999; Rayner et al. 2006). The data were downloaded from UEA on 13 April 2010. Results based on these data were compared to similar results from temperature time series produced at the National Oceanic and Atmospheric Administration (NOAA) National Climatic Data Center (NCDC; Peterson and Vose 1997; Quayle et al. 1999; Smith et al. 2008) and the National Aeronautics and Space Administration (NASA) Goddard Institute for Space Studies (GISS; Hansen and Lebedeff 1987; Hansen et al. 1996, 1999). Both series extend from 1880 through 2009; see also Solomon et al. (2007) and references therein for additional details of the datasets. Except for slight differences in the magnitude of the trends, with a tendency for GISS to exhibit the smallest trends and HadCRUT3 the largest, the results are virtually identical. Accordingly, the NCDC and GISS data will not be further discussed.

Sources of uncertainty for HadCRUT3 and CRUTEM3 are estimated and summarized in Brohan et al. (2006). The estimated 95% level of total uncertainty for global annual HadCRUT3 anomalies is approximately ±0.1°C prior to 1880, ±0.13°C from 1895 to 1930, and ±0.07°C from 1945 to the present (see Fig. 10 of Brohan et al. 2006).

TRENDS OVER GLOBAL SURFACE. Figure 1 shows the well-known evolution of global (HadCRUT3) and land-only (CRUTEM3) annual surface temperature anomalies. A clear upward trend is evident in these series, both for the entire record and for many shorter segments. Large interannual-to-decadal variability is also apparent. These substantial short-term variations can lead to marked differences in trend estimates for time intervals (segments) whose starting and ending dates differ by only a few years.

Temporal variations in global temperature trends are illustrated with the help of two-dimensional parameter diagrams. Figure 2 displays every possible trend (except for year-to-year changes) calculated using a linear regression by least-squares fit. For example, the value plotted at point x = 1980, y = 60 corresponds to the 60-yr trend (°C yr^{-1}) ending in 1980. It is evident that time segments of a few decades or shorter can exhibit either warming or cooling trends, while trends for longer segments are mostly positive, though quite weak compared to those present in shorter segments.

Figure 3a is an alternative presentation showing the cumulative temperature change for each time interval (hereafter referred to simply as the change), as estimated from the fitted linear trend. This representation deemphasizes large but short-lived trends while highlighting sustained long-term trends. Figure 3b shows the same calculation but focusing on the more recent period 1945–2009.

Clearly, positive changes dominate the longer segments. The largest changes are positive and occur for segments longer than 30 years ending in recent years. The maximum change (0.82°C) is observed for the 108-yr period ending in 2009. There is also a

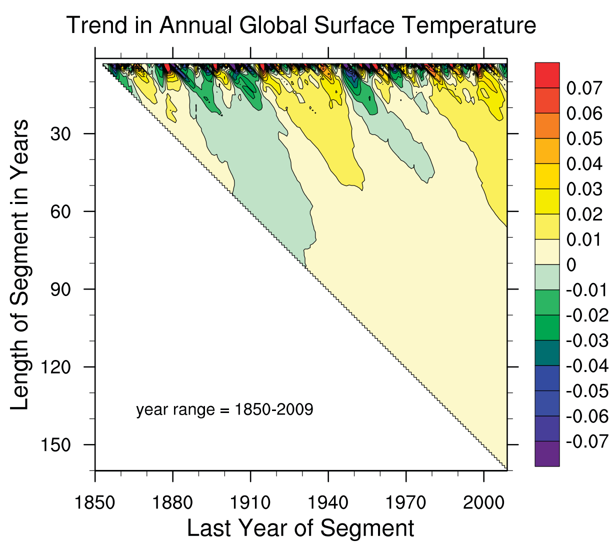

Fig. 2. Trend of HadCRUT3 global annual surface temperatures (°C yr^{-1}) as a function of length of segment and ending year of calculation. Year-to-year changes are not plotted.

secondary peak in warming that took place from the early twentieth century to the mid-1940s, culminating in a 0.62°C rise for the 39-yr period ending in 1945. For the subperiod 1945–2009 (Fig. 3b) every trend longer than 22 years is positive, while for the entire record all segments longer than 82 years exhibit a positive trend.

The overall warming trend, however, is interrupted by brief periods of cooling (e.g., Easterling and Wehner 2009). From 1900 onwards, these cooling periods have not lasted more than 19 years, with two exceptions. The most pronounced episode of extended cooling took place following the warm interlude of 1937–45. When these warm years occur near the beginning of a segment, their influence on the trends extends all the way into the 1980s (Fig. 3a). There is some question, however, as to the accuracy of sea surface temperatures in the middle 1940s (e.g., Thompson et al. 2008). A second, shorter period of cooling took place after the high temperatures of the late 1950s and early 1960s (Fig. 3b).

The recent cooling that has been the subject of much popular media attention (e.g., *Investor's Business Daily*, 4 November 2008) is presently of 9 years duration and amounts to a change of −0.07°C (the 4–8-yr changes ending in 2009 are also all negative). In total, there are 98 positive and 54 negative 9-yr segments. The segment ending in 2009 is only the 44th most negative of the record, well within the range of historical variability. Short-term trends of such magnitude (of either sign), therefore, are far from unusual.

The maximum positive and negative changes for each segment length are shown in Fig. 4. Positive

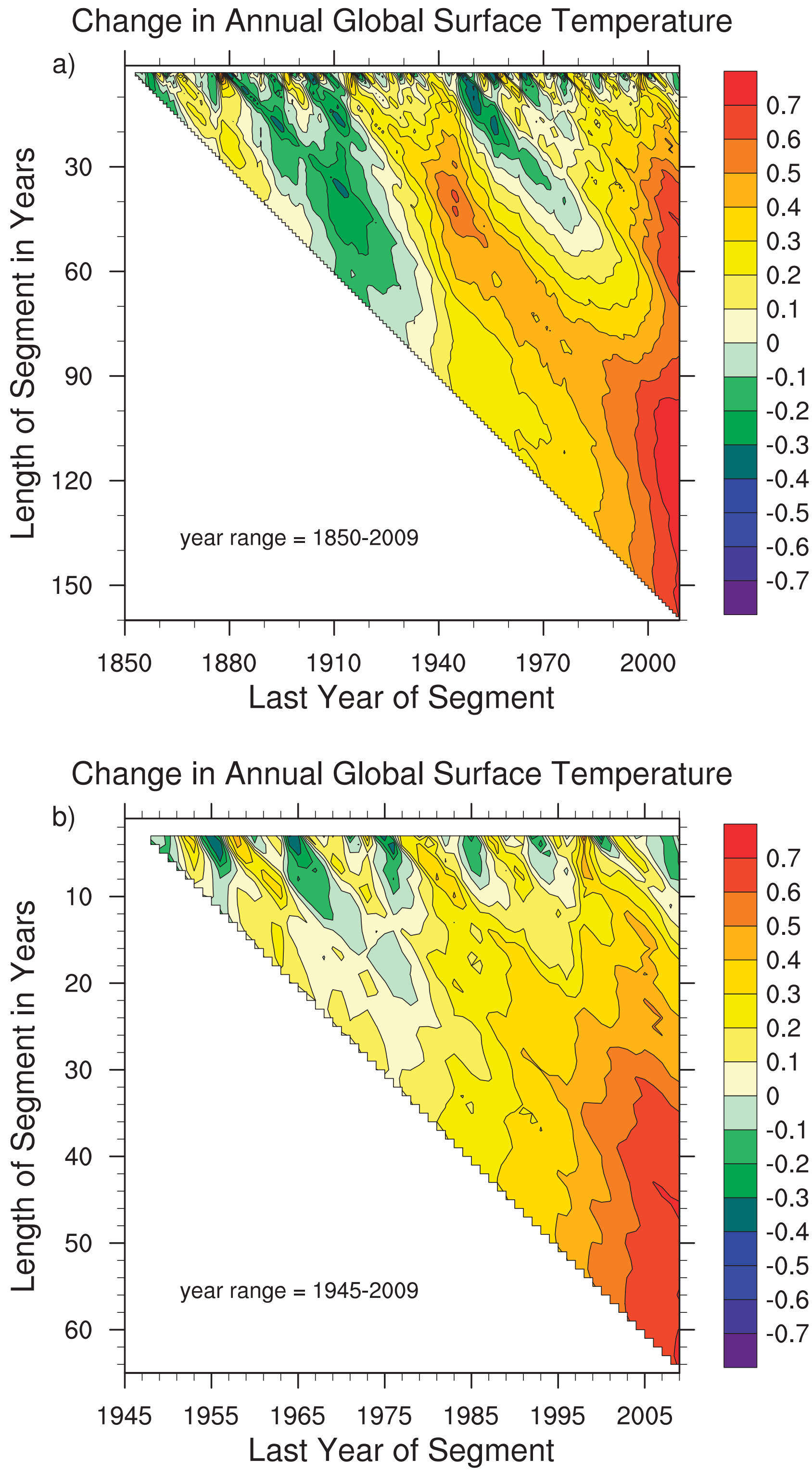

Fig. 3. Change in annual global surface temperature as a function of length of segment and last year of segment. Change is defined as the trend (°C yr^{-1}) multiplied by length of segment. Ranges of analyses are (a) 1850–2009 and (b) 1945–2009.

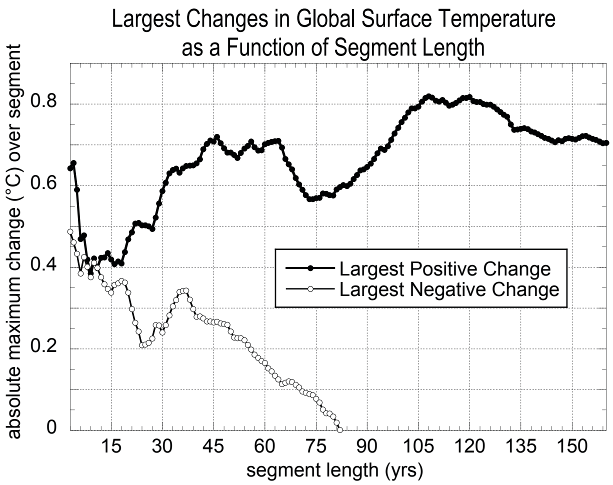

Fig. 4. Largest absolute positive and negative change of global annual average surface temperature as a function of segment length. For each indicated length, all possible subsets of the historical record with the same number of years are compared.

changes are always larger in magnitude than negative changes. The maximum negative change decreases monotonically for segments longer than 36 years. For segments longer than 55 years, the maximum positive change almost invariably corresponds to the segment ending in 2009 (with only eight exceptions).

ESTIMATION OF THE SIGNIFICANCE OF THE TRENDS. There is no commonly agreed optimal method for assessing the statistical significance of trends when the data are strongly autocorrelated and so three different approaches are used here to estimate this significance. The lag-1 autocorrelation of the HadCRUT3 series is 0.9, and the residual serial correlation is 0.74 when the linear trend over the entire record is removed. The Durbin–Watson statistic (Durbin and Watson 1971) confirms that the residual correlation must be accounted for. All methods considered here correct for serial correlation by using autocorrelations and variance for the entire detrended time series. These quantities, however, may not represent the true values over longer periods, which casts some uncertainty onto the results.

The first test is a Student's *t* test (e.g., Wilks 2006). This test measures the probability of detecting a trend of this magnitude if the data were drawn from a random sample. The correction for serial correlation is to assume a first-order autoregressive [AR(1)] process and reduce the number of degrees of freedom accordingly (e.g., Wilks 2006). Changes for segments longer than around 90 years ending after approximately 1985 all exceed the 99.9% level of significance, using a two-sided test (Fig. 5a). Note that this value simply indicates that the null hypothesis (that there is no real trend) can be rejected at that confidence level. Changes longer than around 30 years ending after about 2000 exceed the 97.5% level.

The second test, Fig. 5b, compares trend magnitudes with those in one million randomized samples of the detrended series (a "Monte Carlo" approach). The moving-block bootstrap method (e.g., Wilks 1997, 2006) is utilized to account for serial correlation by randomly resampling by blocks rather than by individual years (with replacement). Optimal sample block lengths vary from 16.7 for the full 160-yr segment to 1.63 for a 3-yr segment and are also computed

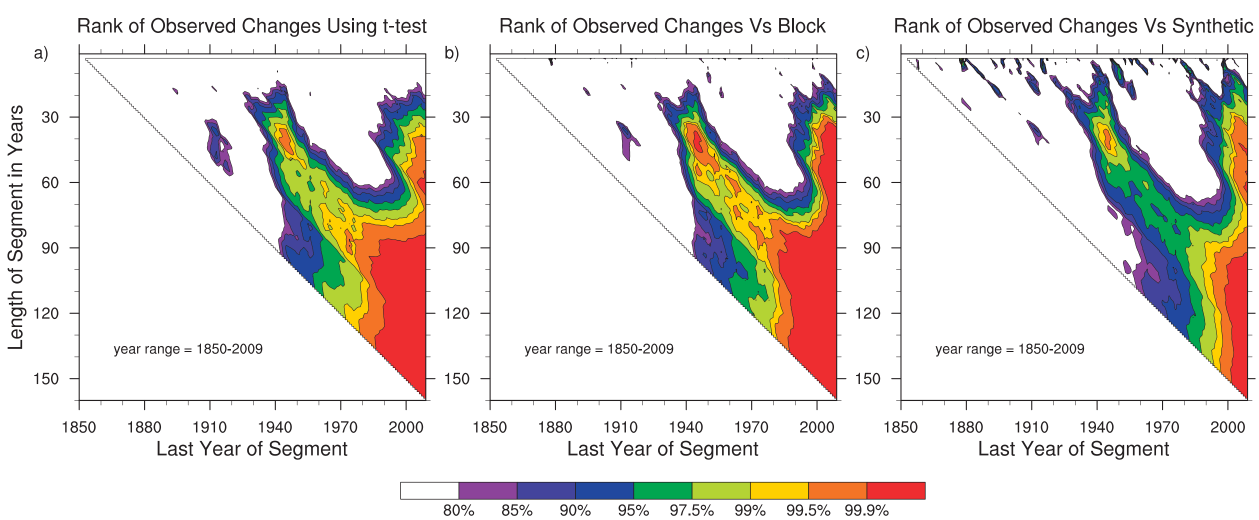

Fig. 5. Percent ranking of absolute value of change in the observed time series compared to (a) a two-sided Student's *t* test; (b) absolute value of change in one million randomized series of same length using the block method; and (c) absolute value of change in one million randomized series of same length using the synthetic series method.

on the assumption of an AR(1) process model. The rankings are qualitatively similar to, but slightly higher than, those from the *t* test (Fig. 5a). Again, very high levels of significance are noted for longer segments ending in recent years. There is also a tongue of high significance for segments longer than about 30 years that begin in the early 1900s.

The first two tests assume an AR(1) process and thus may produce unrealistically high levels of significance, as the autocorrelation functions of both the original and detrended temperature time series exhibit a slowly decaying (rather than exponential) structure (e.g., Wigley et al. 1998). To account for this long-term persistence, the third test compares the observed trends with those in one million randomized synthetic series having the same autocorrelation and variance as the detrended observed series (e.g., Wilks 2006; Box et al. 2008), assuming an AR(10) process model (Fig. 5c). The observed autocorrelation drops below e^{-1} at lag 10. The rankings obtained using this improved, though still imperfect, model are lower than but qualitatively similar to those obtained with the other tests. In particular, the changes in longer segments (>90 years) ending recently still exceed the 99.9% level of significance.

TRENDS OVER LAND ONLY. Figure 6a is similar to Fig. 3a but for continental surface annual temperatures. The recent long-term warming changes over land are larger than the corresponding global changes, with expanding urban heat island effects accounting for a small-to-negligible contribution to this difference (e.g., Jones et al. 1990; Parker 2006). For example, the largest observed land temperature changes reach 0.99°C and occur for the 46-, 126-, and 127-yr segments ending in 2009, compared to 0.72°, 0.79°, and 0.79°C, respectively, for global temperature. Several other segments exhibit changes larger than 0.9°C. All except two of the land changes lasting longer than 53 years are largest for the segment ending in 2009.

In contrast to the peak observed for longer segments ending recently, the secondary warming peak in global surface temperatures observed for ~40-yr segments ending in the 1940s (Fig. 3a) is less evident in the continental temperature record. For instance, the global change for the 39-yr segment ending in 1945 is 0.62°C, whereas the corresponding continental change is only 0.48°C.

As expected, the lag-1 autocorrelation of the detrended continental temperature is lower than for global mean temperature: 0.42 for the entire period (0.54 with the trend included). Figure 6b thus shows the ranking of land-only temperature changes compared to synthetic processes of order AR(8). Most of the land-only segments ranked higher than 80% exhibit rankings similar to those of the corresponding global segments (Fig. 5c), although rankings for the ~40-yr segments ending in the 1940s are lower. Many segments longer than 100 years and ending recently exceed the 99.9% level.

SUMMARY. Two-dimensional parameter diagrams are used to examine time-varying trends in annually averaged global surface and continental temperatures for the period 1850–2009. Every possible trend (longer than two years) and its associated linear temperature change are calculated for the available record.

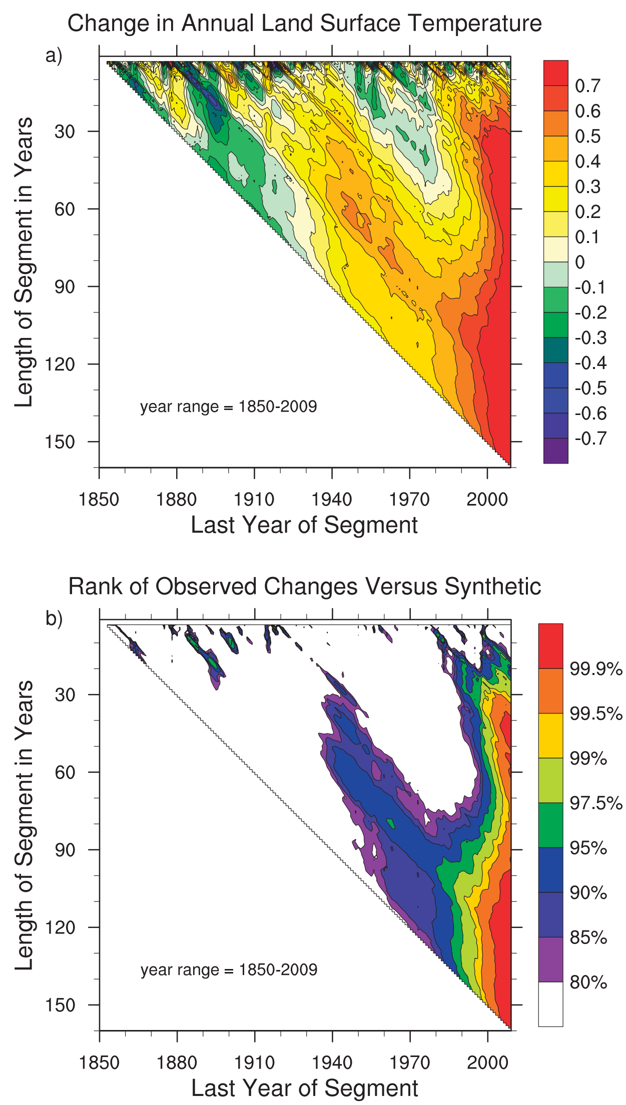

FIG. 6. (a) As in Fig. 3a, except quantity plotted is change in annual land surface temperature. (b) As in Fig. 5c, except quantity ranked is change in land surface temperature.

Changes for segments longer than 82 years have all been positive. Within the constraints of the statistical significance tests, the positive changes of long duration (several decades and longer) ending in recent years are determined to be extremely unlikely to have occurred by chance. A secondary transient peak reflecting warming over a roughly 40-yr period ending in the 1940s is also unlikely to have occurred by chance. Since 1945, all periods longer than 22 years indicate warming, although only those segments ending recently stand out significantly from the noise. The land surface exhibits larger long-term changes than the entire globe, but the transient warming ending in the 1940s is less pronounced.

In contrast, changes shorter than a few decades can be either positive or negative. The recent cooling trend is evident in the global record beginning in 2001. Such changes, however, are not statistically significant and are in fact relatively common in the historical record.

Conclusions from the simple models used here to determine statistical significance must be tempered by several caveats. In addition to the uncertainties in the statistics used to model the observed temperature behavior, the assumed models may not reflect the true behavior of the complex atmosphere. The determination of statistical significance depends largely on the assumed statistical process and parameters (e.g., Woodward and Gray 1993; Cohn and Lins 2005). For example, had the effective number of degrees of freedom for the *t* test of Fig. 5a been calculated from the original rather than the detrended series, the significance levels would have been lower than those for the other tests (Figs. 5b and 5c). Another, perhaps better, approach to determining whether or not the observed trends are fortuitous is to seek physical attribution mechanisms (Barnett et al. 2005).

The simple method described here has applications beyond the analysis of global surface temperatures. It should be of value in ascertaining the sensitivity of trends to choices in the start and end points of the time series and assessing whether particular periods are representative (or not) of longer-term trends.

ACKNOWLEDGMENTS. The authors wish to thank Joe Barsugli, Chris Folland, John Kennedy, Prashant Sardeshmukh, Huug van den Dool, and an anonymous reviewer for their considerable effort that resulted in extremely worthwhile suggestions for an improved manuscript. The encouragement from Robin Webb and Marty Hoerling is also appreciated.

REFERENCES

Barnett, T., and Coauthors, 2005: Detecting and attributing external influences on the climate system: A review of recent advances. *J. Climate,* **18,** 1291–1314.

Box, E. P., G. M. Jenkins, and G. C. Reinsel, 2008: *Time Series Analysis: Forecasting and Control.* John Wiley and Sons, 746 pp.

Brohan, P., J. J. Kennedy, I. Harris, S. F. B. Tett, and P. D. Jones, 2006: Uncertainty estimates in regional and global observed temperature changes: A new dataset from 1850. *J. Geophys. Res.,* **111,** D12106, doi:10.1029/2005JD006548.

Cohn, T. A., and H. F. Lins, 2005: Nature's style: Naturally trendy. *Geophys. Res. Lett.,* **32,** L23402, doi:10.1029/2005GL024476.

Durbin, J., and G. S. Watson, 1971: Testing for serial correlation in least squares regression. III. *Biometrika,* **58,** 1–19.

Easterling, D. R., and M. F. Wehner, 2009: Is the climate warming or cooling? *Geophys. Res. Lett.,* **36,** L08706, doi:10.1029/2009GL037810.

Hansen, J., and S. Lebedeff, 1987: Global trends of measured surface air temperature. *J. Geophys. Res.,* **92,** 13 345–13 372.

——, R. Ruedy, M. Sato, and R. Reynolds, 1996: Global surface air temperature in 1995: Return to pre-Pinatubo level. *Geophys. Res. Lett.,* **23,** 1665–1668.

——, ——, J. Glascoe, and M. Sato, 1999: GISS analysis of surface temperature change. *J. Geophys. Res.,* **104,** 30 997–31 022.

Jones, P. D., 1994: Hemispheric surface air temperature variations: A reanalysis and an update to 1993. *J. Climate,* **7,** 1794–1802.

——, and A. Moberg, 2003: Hemispheric and large-scale surface air temperature variations: An extensive revision and an update to 2001. *J. Climate,* **16,** 206–223.

——, P. Ya. Groisman, M. Coughlan, N. Plummer, W.-C. Wang, and T. R. Karl, 1990: Assessment of urbanization effects in time series of surface air temperature over land. *Nature,* **347,** 169–172, doi:10.1038/347169a0.

——, M. New, D. E. Parker, S. Martin, and I. G. Rigor, 1999: Surface air temperature and its variations over the last 150 years. *Rev. Geophys.,* **37,** 173–199.

Parker, D. E., 2006: A demonstration that large-scale warming is not urban. *J. Climate,* **19,** 2882–2895.

Peterson, T. C., and R. S. Vose, 1997: An overview of the Global Historical Climatology Network temperature database. *Bull. Amer. Meteor. Soc.,* **78,** 2837–2849.

Quayle, R. G., T. C. Peterson, A. N. Basist, and C. S. Godfrey, 1999: An operational near-real-time global temperature index. *Geophys. Res. Lett.,* **26,** 333–335.

Rayner, N. A., P. Brohan, D. E. Parker, C. K. Folland, J. J. Kennedy, M. Vanicek, T. Ansell, and S. F. B. Tett, 2006: Improved analyses of changes and uncertainties in sea surface temperature measured in situ since the mid-nineteenth century: The HadSST2 dataset. *J. Climate,* **19,** 446–469.

Smith, T. M., R. W. Reynolds, T. C. Peterson, and J. Lawrimore, 2008: Improvements to NOAA's historical merged land–ocean surface temperature analysis (1880–2006). *J. Climate,* **21,** 2283–2296.

Solomon, S., and Coauthors, Eds., 2007: *Climate Change 2007: The Physical Science Basis.* Cambridge University Press, 996 pp.

Thompson, D. W. J., J. J. Kennedy, J. M. Wallace, and P. D. Jones, 2008: A large discontinuity in the mid-twentieth century in observed global-mean surface temperature. *Nature,* **453,** 646–649, doi:10.1038/nature06982.

Wigley, T. M. L, R. L. Smith, and B. D. Santer, 1998: Anthropogenic influence on the autocorrelation structure of hemispheric-mean temperatures. *Science,* **282,** 1676–1679, doi:10.1126/science.282.5394.1676.

Wilks, D. S., 1997: Resampling hypothesis tests for autocorrelated fields. *J. Climate,* **10,** 65–82.

——, 2006: *Statistical Methods in the Atmospheric Sciences.* Academic Press, 627 pp.

Woodward, W. A., and H. L. Gray, 1993: Global warming and the problem of testing for trend in time series data. *J. Climate,* **6,** 953–962.

AMS MEMBERS

GIVE A GREAT GIFT AT A GREAT PRICE

Looking for the perfect present for the weather enthusiast in your life? Want to make a valuable contribution to your local library or community college? Send a subscription to *Weatherwise* magazine (6 issues) for just $24.95*—That's nearly 50% off the list price!

Written for a general audience, Weatherwise offers a colorful and nontechnical look at recent discoveries in meteorology and climatology. Check out the latest table of contents at www.weatherwise.org.

Want your own? Then order a personal subscription at the same great price.

Use your 2011 dues renewal to place all of your *Weatherwise* orders today!

*Cost for delivery outside of the U.S. is $40.95.
Weatherwise is available to AMS Members through a cooperative agreement with Taylor & Francis Group LLC, the publishers of *Weatherwise*.